DIW Berlin

German Institute
for Economic Research

Discussion Papers

486

Richard V. Burkhauser*
Dean R. Lillard*

The Contribution and Potential of Data Harmonization for Cross-National Comparative Research

Berlin, April 2005

* Cornell University and DIW Berlin

IMPRESSUM

DIW Berlin
Deutsches Institut für Wirtschaftsforschung
Königin-Luise-Str. 5
14195 Berlin
Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
www.diw.de

ISSN print edition 1433-0210
ISSN electronic edition 1619-4535

The Contribution and Potential of Data Harmonization for Cross-National Comparative Research

Richard V. Burkhauser and Dean R. Lillard
Cornell University and DIW Berlin

Corresponding author:
Dean R. Lillard
Department of Policy Analysis and Management
247 MVR Hall
Cornell University
Ithaca, NY 14853-4401
USA

Tel. (607) 255-9290
FAX (607) 255-4071
E-mail DRL3@cornell.edu

Acknowledgements
We thank Joachim Frick and Robert Walker for comments. An earlier version of this paper was presented at the 2004 fall conference of the Association for Public Policy and Management in Atlanta, GA.

Abstract

The promise of empirical evidence to inform policy makers about their population's health, wealth, employment and economic well being has propelled governments to invest in the harmonization of country specific micro data over the last 25 years. We review the major data harmonization projects launched over this period. These projects include the Luxembourg Income Study (LIS), the Cross-National Equivalent File (CNEF), the Consortium of Household Panels for European Socio-Economic Research (CHER), the European Community Household Panel (ECHP), the European Union Statistics on Income and Living Conditions (EU-SILC), and the Survey of Health, Aging and Retirement in Europe (SHARE). We discuss their success in providing reliable data for policy analysis and how they are being used to answer policy questions. While there have been some notable failures, on the whole these harmonization efforts have proven to be of major value to the research community and to policy makers.

Introduction

Most economically developed countries regularly survey a large representative sample of their populations**.** They do so to document the economic well being, labor market outcomes, and health of their citizens and to gauge whether and how effectively public policies have improved their lives, as measured by social success parameters developed from the data. In addition, several countries have mature longitudinal surveys that attempt to capture movement in these measures as well as the life course events that influence them from a dynamic perspective.[1] In addition to these representative surveys of the entire age distribution, some countries field longitudinal micro-samples that focus on different age cohorts as they pass through some critical life event.[2]

While most of these studies were developed to evaluate outcomes in a specific country, over the last 25 years these data have been increasingly used in cross national comparative studies (Burkhauser and Smeeding, 2001). Here we describe the promises the research community has made to encourage public policy makers to fund both the harmonization of existing country data and the creation of more comparable country data to facilitate cross-national research. We focus on the major data projects that have been launched over this period and broadly survey the social science based research they have generated. Our purpose is to document the degree to which these harmonization efforts have succeeded and the extent to which researchers have used these data to answer policy questions. While there have been some notable failures, on the whole these data efforts have proven to be of major value to the research community and to policy makers.

---

1 Examples include the US Panel Study of Income Dynamics, the German Socio-Economic Panel, the British Household Panel Study, the Malaysian Family Life Surveys, and the Russian Longitudinal Monitoring Study.
2 Examples include the National Child Development Study in Great Britain, the Health and Retirement Survey in the United States, and the English Longitudinal Study on Aging.

What Policymakers Want to Know

Most public policy analysts aim to provide policy makers with carefully researched evidence that will inform and shape the policies that are developed and implemented. For their part, policy makers have increasingly recognized that firmly grounded evidence promises to improve the policy making process. That recognition has helped convince policy makers to invest in projects over the last 25 years that proposed to first harmonize country specific data and then to use those data to answer policy questions. These questions are of three general types.

What is happening to outcomes of interest? Answers to this type of question constitute the first and most basic information researchers provide policy makers. Large representative cross-sectional samples of policy relevant populations provide regular snap-shots of the labor earnings, economic well-being, health, and work behavior of those populations and of vulnerable subgroups within them. Such data allow researchers to answer questions like, "Is average income going up or down?" or "Is the share of income earned by the poorest 20 percent of households greater now than it was ten years ago?" Social success indicators derived from these data—e.g., average wage rate, average income, poverty rates, wage and income dispersion, and unemployment rates—provide quantitative evidence of economic and social policy outcomes.

By contrast, panel data allow policy analysts to not only describe outcomes at a point in time, they also allow one to describe how those outcomes are evolving over time. For example, with panel data, researchers can document the wages of a worker in a given period and how those wages evolve over the worker's lifetime. The same is true for panel data that measure household income or poverty. The information provided by panel data yields additional insights. While cross-sectional data allow researchers to document the number of persons or household who have low income or are in poverty – important information for policy decisions, panel data

additionally allow researchers to document the length of time persons or households experience those periods of deprivation. More recently, as health and the impact of social policies on health have come to the forefront as a public policy concern, studies collecting both cross-sectional and panel data have included measures of health. These new data allow analysts to document and describe inequality in health outcomes, across populations and over time. Researchers have developed both social success measures from these data and models that describe their levels, trends, and risks of occurrence across the life cycle.

Why are given outcomes observed? Social science researchers have used these data to disentangle the impact of socio-economic characteristics, market forces, and past public policies on individual economic well-being, labor market outcomes, and health. The data allow policy analysts to answer questions like, "Does unemployment rise when legislation reduces the legal hours of work?" or "How does women's labor force participation change when anti-discrimination laws are passed and enforced?" Researchers have answered questions like these from both a static and dynamic perspective using behavioral models in economics and sociology.

How can "bad" outcomes be reduced and "good" outcomes encouraged? Public policy researchers have begun to use these data to predict the behavioral consequences of future public policies using two types of models. The first type of model describes in detail the structure of decisions leading to an outcome of interest (structural models). The second type of model takes advantage of changes in policies that occurred outside the control of persons being studied (so-called natural experiments). These two types of models allow researchers to study how actual changes in policy affect an outcome of interest without having to conduct social experiments in which participants are randomly assigned one policy or another. For example, if two countries paid different public pension amounts to people of the same age, one can study how retirement

decisions in each country varied with the generosity of retirement income. Further, if public pensions within a country vary across persons born in different years, one can use differences in pension benefits both within and across countries to identify how changes in pension benefits affect retirement decisions.

The Value Cross-National Research Adds

Cross-national comparative research contributes basic knowledge and improves our understanding of how policies affect choices people make. Cross-national comparative research deepens our understanding of human behavior because it helps researchers identify behavior that is common across different cultures and societies. For example, cross-national research shows that retirement decisions in every country vary systematically with the generosity of income workers can expect in retirement – as determined by public and private pension plans (see Gruber and Wise, 1999). Such evidence points to a basic relationship between retirement income and decisions to exit from the labor force that is a fundamental human behavior not something that is tied to a particular society or culture. Cross-national comparative research also improves our understanding of how policies affect choices people make because it affords researchers the opportunity to study the effects of a much broader and richer mix of policies than typically available in a single country. For example, Kenkel, Lillard and Mathios (2004) document the substantial variation across the U.S., U.K., Germany, and the Russian Federation in the mix of taxes, workplace smoking bans, advertising restrictions, and warning label policies aimed at reducing consumption of tobacco. They note that there is as much or more variation across countries than within countries in the different mixes of these policies and that these differences can be used to study how decisions to smoke vary with policy mixes outside the

range observed in a single country. Thus, cross-national comparative research adds value to our understanding of basic human behavior and to the analysis of how public policies affect choices.

Two Traditions of Micro Data Development

Cross national comparative analysis is supported by two basic forms of micro data. The first is the all purpose micro-sample. This type of data collection can either be a regularly fielded cross-sectional survey or a panel survey. The *Current Population Survey* (CPS) in the United States as well as the *Family Expenditure Surveys* (FES) and the *Family Resources Survey* (FRS) in the United Kingdom are excellent examples of regularly fielded cross-sectional surveys. The CPS, FES, and FRS are random samples of each country's entire population.

While these cross-sectional surveys collect information on the demographic characteristics of members of the sampled households, their primary aim is to collect data on income, employment, and, in the case of the UK studies, consumption expenditures. Consequently, both surveys contain little information on health or wealth. Most OECD countries regularly field these types of surveys.

Many countries also have panel surveys whose core questions yield similar data on demographics, income, and employment. The first and longest running socio-economic based panel is the United States Panel Study of Income Dynamics (PSID). But the PSID has inspired a generation of similarly focused panels in other countries over the last 20 years. Many of these panel surveys periodically include special modules that gather information on a variety of topical issues such as: health, wealth, and family history. Because panel surveys re-interview the same respondents, these surveys contain a richer mix of socio-economic information than the data gathered in a single cross-sectional survey. In recent years, many have incorporated health questions from their special topic modules into their set of core questions.

An important new source of data for cross-national research has been the creation of panel surveys capturing the economic well-being, labor force outcomes, health and wealth of a cohort of older working-age people. The earliest such cohort survey, the United States *Retirement History Survey*, funded by the Social Security Administration, followed men and unmarried women for ten years; from 1969, when they were aged 58 to 63, until 1979. But the model for almost every recent OECD country cohort study is the *Health and Retirement Study* (HRS) in the United States. The HRS was funded by a consortium of government agencies led by the National Institute on Aging. Since 1992 the HRS has followed a cohort of men and women aged 51 to 61. Additional cohorts of men and women have been added since then. The *English Longitudinal Study in Aging* (ELSA) owes much of its design to the HRS. ELSA began following a cohort of men and women aged 50 and over in 2002. The first wave of ELSA data has been released and a second round of data collection was completed in 2004. But the newest and most ambitious cohort study is the multi-country *Survey of Health, Aging and Retirement in Europe* (SHARE) which released its first wave of data for 10 European countries in 2005.

Cross-National Comparisons of Comparable Country Data

Cross national research is inherently more difficult to perform than research focused on a single country. It is more difficult to create comparable data across countries than to simply focus on a single country. Even when comparable data are available, the institutional differences across countries often require a team of country specific experts to appropriately use these data.

The first of these stumbling blocks to cross national research has been overcome by two types of research efforts. The first type has drawn data from existing single country data sets and ex post harmonized it to facilitate cross-national comparisons. The second type of research effort has, ex ante developed common survey instruments or created new data sets that are intended to

be comparable. While this latter method is conceptually appealing, it has proven to be difficult to implement in practice.

Concepts of Harmonization

The goal of data harmonization is to create data that measure the same conceptual variable and that are measured in the same units. In some cases, data can be easily harmonized across studies. For example, given information on the month and year an individual was born, it is straightforward in both concept and practice to define age in common units of years. For variables other than age and sex, the task of data harmonization quickly grows more complex. The harmonization of variables such as "marital status" typically requires that a researcher collapse multiple categories in one data set to the number of categories available in the data set that has the fewest categories defined. The most complex types of harmonization efforts mounted to date often involve income. The challenge of data harmonization of income is to first count income received over a similar time period (e.g. a calendar year). More importantly, harmonization efforts need to account for widely different tax schemes present in different countries. For example, in the Cross-National Equivalent File (CNEF) intensive efforts have been made to estimate for each country in the CNEF, the amount of taxes households must pay out of household income in each year. In addition, the CNEF sums all transfer income households receive from government and private sources tabulated in each survey. The harmonized income data in the CNEF therefore measure household income after taxes have been subtracted and after transfers have been made (see Burkhauser et al., 2001). Ultimately the goal of harmonization is to create variables measuring conceptually equivalent data that can be directly and easily compared.

Research Strategies with Harmonized Data

Even when harmonized cross-country data are available, researchers face the difficult problem of how to use the data effectively. Some researchers have used the strategy of gathering researchers most involved in the creation of the individual data into a team to do research on a common topic (Burkhauser et al., 1991; Duncan et al., 1993; Jenkins et al., 2003; Börsch-Supan et al., 2005; Burkhauser et al., forthcoming). Others have organized a team of researchers from various countries who have been doing similar research on country-specific samples. The research team then agrees on a common method to investigate a common problem and searches for the country data necessary to answer the question (Haveman et al., 1984; Aarts et al., 1996; Gruber and Wise, 1999, 2004).

Successes and Failures on the Road to Harmonization

To frame our discussion of successes and failures, we tabulate currently available harmonized data efforts into three broad categories. These three categories are ex post harmonization of existing cross-sectional data, ex post harmonization of existing panel data, and ex ante harmonization of new panel data. We then assign each data harmonization effort to the category that best describes how the harmonization was done. Finally, we identify whether each harmonized data set can be used to address four major substantive area of policy. The four broad areas we consider include income, labor outcomes, health, and wealth. As our discussion of Table 1 will show, harmonization efforts have succeeded in providing researchers with excellent data in some of these substantive policy areas. This research has, in turn, provided empirical evidence to policy makers in some of these areas. But much remains to be done. The appendix gives more details (including internet addresses) of the major data harmonization efforts discussed in this paper.

Existing Cross-Sectional Data. The earliest and most successful attempt at harmonizing existing data sets is the *Luxembourg Income Study* (LIS). The frame and types of data available in each country sample varies but most country samples are similar in design to the United States CPS and contain excellent information on household income and labor market outcomes. Since 1980 LIS has worked with the statistical ministries of many countries to store their cross-sectional micro-samples in the LIS home office. Researchers are allowed to write programs that are implemented by LIS staff in Luxembourg. While this system allows researchers to use the data, it does not permit them to see the original data or to challenge the assumptions and methods used by LIS personnel to harmonize the data.

Nonetheless, as the extensive use of LIS in several policy literatures attests, LIS has been extremely successful in facilitating cross-national comparisons of the major industrial countries. LIS country data has been a major asset, for example, in the growing literature on cross national comparisons of poverty and inequality in OECD countries (Smeeding, 2004; Gottschalk and Smeeding, 1997.)

Existing Panel Data. Two ongoing harmonization efforts have been carried out with respect to mature panel data and each is based on the LIS model. The first, the *Cross-National Equivalent File* (CNEF), harmonizes a subset of the data found on four mature panel data sets: The United States *Panel Study of Income Dynamics* (PSID), The German *Socio-Economic Panel* (SOEP), The British *Household Panel Study* (BHPS) and the Canadian *Survey of Labor and Income Dynamics* (SLID). CNEF primarily contains information on income and labor market outcomes but recently added health variable as each of its country data sets began to add a richer mix of health variables to their core questions. CNEF uses the PSID as its model and harmonizes its key variables to the definition of variables in the PSID. By doing so, it provides a data set that

is especially useful for making comparisons between outcomes in the United States to those in the other three countries. It has the advantage over LIS of not only allowing researchers access to the original data sets from which the CNEF variables are created but also to the programs used to create them. Access to these programs allows individual researchers to review the algorithms used to create variables. It also allows researchers to customize the programs. Efforts are made to make it easy for researchers to merge CNEF data with data from each parent study. In this way researchers can append information from the original data to create new harmonized variables that are then made available to the cross-national research community.

A second harmonization effort, the *Consortium of Household Panels for European Socio-Economic Research* (CHER), focuses primarily on panel studies in European countries. The CHER began as a feasibility study for a data production and dissemination exercise. It developed and enhanced a comparative database for longitudinal household studies by harmonizing and integrating micro datasets from a large variety of independent national panels and from the *European Community Household Panel* (ECHP). The current database, which is available for comparative research, contains data from 1990 to 2001 on 18 European countries. Like CNEF, CHER uses existing panel data. While in many ways the CHER harmonization effort is similar to that of CNEF, it differs on two important dimensions. First, it attempts to harmonize data across mostly European countries. More importantly the project uses the variable definitions in the German SOEP as their model. These data are especially useful for making comparisons between European countries.[3]

3A third harmonization effort that is no longer being updated is the EPAG (*European Panel Analysis Group)* dataset. It includes data from the BHPS, SOEP, and the *Socio-Economic Panel* (SEP) of the Netherlands. EPAG is also modeled on the SOEP and like the CNEF and CHER, primarily harmonizes data on income and employment (see http://www.irc.essex.ac.uk/epag/dataset.php).

The effort to harmonize existing panel studies share one significant organizational feature: active researchers conceived, planned, and carried out how the data would be harmonized. While data managers, some in government statistical agencies, were often involved in the process, it was researchers who decided how to define equivalently the variables of interest. In addition, the above efforts have involved researchers familiar with the institutions of each country. This involvement means that, when a decision had to be made about how to harmonize data, the decision was not only informed by country-specific knowledge of institutions but also was guided by an overall conceptual definition based on the latest research on that specific topic. Even using similarly designed country panel surveys, it is not a trivial exercise to harmonize the data consistently across countries. Researchers guided by theory and concepts flowing from the research pertinent to the object of their studies are best able to make the assumptions necessary to harmonize data across countries.

While less cross-national research has been done using these harmonized pre-existing panels than is the case for LIS, these data are increasingly being recognized in the policy literature. This recognition occurs most in areas of public policy that focus on how outcomes evolve over the life course. For example, in a chapter in the OECD publication *Economic Outlook* (2001) that examines the dynamic nature of poverty, the authors' state:

"The CNEF data are extremely valuable for providing long panels that enable more comprehensive and detailed analysis of poverty dynamics, both for pre- and post-fiscal income. These data enable comparisons of the effects of national tax and transfer systems by providing the appropriate income variables defined identically." (Economic Outlook, 2001).

Chapter 2 of *Economic Outlook* (2002), *Women at Work: Who are they and how are they faring* also make use of CNEF data in capturing employment transitions by sex (long term

duration) and continuity in employment status by education and children present. Such cross-national comparisons of dynamic labor market measures were not possible before CNEF harmonized these data.

Most of the early researchers who helped to develop the first generation of panels, and who were major forces in the harmonization of these panels, primarily focused their research on questions of economic well-being. Hence, it is not surprising that income measures make up a large part of these panels' cores. It is also not surprising that the first major cross-national project that utilized panel data focused on poverty dynamics in eight countries (Duncan et al., 1993). Greg Duncan organized researchers from eight countries each of whom had access to either self reported panel data, or administrative records that enabled them to track the income of their country's population. This team developed common measures of income and poverty and produced the first modern cross national poverty duration measures (Duncan et al., 1993).

This team and numerous other researchers have used harmonized panel data to capture the dynamic patterns of poverty, both within and across countries. They have also provided a life course picture of poverty that is remarkably similar across countries. Using harmonized panel data, researchers have shown that:

- Mobility into and out of poverty is high in every nation
- The "permanent" underclass is small relative to the entire population who experience a spell of poverty
- Poverty touches many more people over their life course than is apparent from a one-year snapshot.

Partly as a consequence of these research findings, policymakers have begun to craft policies that better target resources to the long term poor. (See Burkhauser, 2001 for a fuller discussion.)

New All Purpose Panels. The most comprehensive attempt to create an ex ante harmonized data set is the *European Community Household Panel* (ECHP). Led by Eurostat, the ECHP attempted, by using a common survey instrument, to create a set of country based data sets that were comparable across countries. The ECHP goal was to create comparable panel data for all European Union (EU) countries.

While these data were collected from 1994 through 2001, the ECHP's goal of creating harmonized data through a common survey instrument was not successful. The panels were abandoned in 2001. The ECHP was plagued by problems from the outset. In part these problems may have arisen because the ECHP was developed by Eurostat and implemented by each country's statistical agency with little or no consultation with the research community. Hence, unlike the successful harmonization efforts discussed above, end users played a minor role in the creation and implementation of the survey instrument. Most troubling, the ECHP project failed to utilize the long experience of researchers who were running mature panel surveys in EU countries.

After two waves of ECHP it was clear that several key country panel surveys had unsustainable attrition problems. In Germany, Great Britain and the Netherlands, ECHP panels were abandoned and replaced with existing panels (SOEP, BHPS, and SEP respectively) run by researchers outside their country's statistical agencies.

After collecting data for only eight years, the ECHP ended in 2001. The accumulated problems which led to its demise included:

- Long delays in processing
- Problems with initial responses
- Problems with attrition rates

- Non-uniform implementation
- Lack of input from the research community in design and response to users over time
- Initial failure to take advantage of existing panels (SOEP, BHPS, and SEP)
- Poor dissemination strategy to get the data to the international research community
- High costs of use for individual researchers.

Despite these problems, the ECHP may be of some use to the research and policy communities. The consortium of researchers who run the European Panel Analysis Group has devoted considerable resources to making the ECHP available and useful to the research community. They have established a group of users called the European Panel Users Network (EPUNet) who are dedicated to lowering the cost to potential researcher users of the ECHP (see http://epunet.essex.ac.uk/echp.php#introduction). The EPUNet group is also investigating whether, despite the substantial attrition in the ECHP, some of the country panels can provide reliable evidence for policymaking.

The European Union—*Statistics on Income and Living Conditions* (EU-SILC) has been proposed as a replacement for the now discontinued ECHP. EU-SILC will provide comparable and timely cross-sectional data on income, poverty, social exclusion and other living conditions as well as longitudinal data restricted to income, employment and a limited number of non-monetary indicators of social exclusion. Unlike the ECHP, the EU-SILC will be harmonized ex post, so it will avoid the difficulties of ex ante harmonization. It is also far less ambitious in its panel goals. Its longitudinal component is limited to a four-year rotating panel. While this will allow for some dynamic analysis, it will be over a time period that is far shorter than that available in the CNEF panels.

New Cohort Panels. The Health and Retirement Survey (HRS) has been a major source of information on the economic well-being, labor force behavior, health and wealth of men and women transitioning into retirement age in the United States over the last decade. It was the inspiration for the English Longitudinal Study in Aging (ELSA) in England and it is the model for a far more ambious effort to create similar panel data sets in other European countries. The *Survey of Health, Aging and Retirement in Europe* (SHARE) has released its first wave of data (collected in 2004) in 2005. Like ELSA, it consists of a panel of men and women aged 50 and over. SHARE is being lead by a European network of researchers who are working together with United States based researchers at National Bureau of Economic Research (NBER). SHARE's major advantage over previous efforts like ECHP is that it is being led by a team of outstanding country based researchers who have been working with their country's all purpose panel survey data. Hence, this team is well aware of the problems of fielding panel surveys and has already avoided some of the pitfalls that befell the ECHP project. The SHARE researchers consulted with researchers who created the HRS and ELSA surveys in the development of their original set of English language questions. In addition, when they translated questionnaires, they used experts who were not only fluent in each country's language but who were also familiar with each country's social environment. The SHARE researchers have also made the early release of data a major priority and are doing so in a way that imposes a minimal cost (in effort only) on the researcher. While it is far too early to evaluate the success of this panel project, early indications are that it will be the first successful ex ante harmonized data set for Europe.

A New Generation of Research

Retirement. A new generation of researchers has begun to use harmonized cross-national panel data to estimate more structural models of behavior in order to determine how that

behavior will be affected by policy changes. These models are focusing on two areas of behavior that are of growing concern to policy makers and social scientists: retirement and health behavior.

The promise of disentangling the relative importance of health and public policies on retirement was one of the major reasons for the creation of ELSA, as well as for NBER researchers and their European collaborators developing the new SHARE surveys. In addition, the NBER-led research team has produced two major research volumes on the transitions out of the labor force for older workers in 11 OECD countries. Like the Duncan team a decade earlier, this team, headed by Jonathan Gruber and David Wise, focused on a single common problem—the importance of a country's social security system on the timing of retirement. They then developed a common method of estimating that effect and used country experts to carry out the analysis. In their first effort, team members primarily used cross sectional data to perform their analyses (Gruber and Wise, 1999). In their second effort, the NBER research team used more sophisticated structural modeling to estimate the impact of social security on the timing of retirement in each country. They did so using a common estimation method and existing panel data (Gruber and Wise, 2004).

The next phase of this effort is for team members to develop a new generation of data in their home countries based on the HRS and ELSA. Team members are providing the intellectual vision for the new SHARE data. Unlike the ECHP effort, this team of researchers is creating these data independent of their country's statistical agencies and with a well conceived end use to guide them in their efforts to create common variables.

Health. Inequalities in health outcomes have become a major social issue. There is a major dispute in the public policy literature as well as in the public health and economics

literature over the cause of these inequalities. The public health literature argues that not only is health a function of past income but it is also a function of past income inequality. The evidence for this view is mixed. Two major literature reviews suggest that that much more careful modeling is necessary to disentangle these two effects and that better data will be necessary to test these models. (See Wagstaff and van Doorslaer, 2000 and Deaton, 2003).

Better data is on the way. In the last few years the PSID, BHPS and SOEP have added new health variables to their core questions. The addition of these health variables will allow researchers to use these panels to develop models that can test the factors that impact on health at older ages. CNEF will provide a new set of variables on disability and health in its next wave.

Lillard and Burkhauser (2005), using CNEF data, provide an example of the research that cross-national panel health data permit. They measure the independent effects of past income and current income inequality on the current health of men and women at various ages and find little evidence of an independent effect of inequality once income is controlled.

While only the first wave of SHARE is now available to researchers, its rich set of health questions will offer researchers a consistent set of self-reported health variables on 10 European countries that closely follow the health questions asked in HRS and ELSA. If SHARE is able to successfully avoid the problems that plagued the panel aspects of the ill-fated ECHP it will be a major new source of cross national data on the health of older persons.

Discussion and Conclusion

Cross-national research using large representative data sets is still relatively new. It has only been over the last 25 years that harmonized cross-sectional data have become available to the international research community, primary via LIS. And it is only in the last decade that several ex post harmonized country panel data sets have become available. Yet these panel data

have already become essential for those interested in knowing the relative economic well-being of OECD populations and their labor market outcomes. Dynamic cross-national analysis is now common on issues related to income mobility, poverty dynamics, and employment duration.

The new generation of researchers interested in structural modeling will increasingly use panel data to test their models and they are already doing so with respect to exit from the labor market at older ages. Researchers directing established panel data sets have increased their commitment to collecting information on health. In addition, many of the new generation of cohort panel surveys include core questions on health. Both of these data efforts offer the promise that cross-national researchers will have access to the data needed to not only understand the importance of socio-economic variables on health outcomes but also to document and understand how policies influence health behavior and how health is related to and influences major life course events like exit from the labor force.

Many of the promises researchers made to obtain funding for cross-national research and the data sets necessary to conduct it have been kept. As Table 1 shows, excellent harmonized cross-sectional data now are available to track economic well being (as measured by income) and labor outcome in LIS with the promise of a new generation of such data from EU-SILC. CNEF and EPAG provide such data for several ex post harmonized mature panels. HRS and ELSA do so for older cohorts. SHARE promises to spread these comparable cohort data sets to many other European countries over the next decade. Harmonized information on health is harder to find but EU-SILC could be the platform for collecting health information linked to economic variables in harmonized cross-sectional data sets. Furthermore, we are already seeing a much greater commitment on the part of the mature all purpose data sets (PSID, BHPS, SOEP) to add health variables to their core questions and for CNEF to harmonize these data for cross-national

researchers. But most importantly, HRS, ELSA, and SHARE have from their inception made health information central to their data collection plans.

Cross-national information on wealth has also improved. HRS, ELSA and SHARE have been committed to its collection from the start. The PSID began regularly using its special modules in the mid-1990s to collect wealth information and LIS is now involved in a series of efforts to collect information on wealth at the country level. But as the history of the ECHP shows, not all investments in data have produced benefits that exceeded their costs. And while SHARE has the potential to succeed where ECHP failed, much remains to be done before it will be possible to get a set of truly *ex ante* harmonized panel data sets in the hand of researchers. The experience gained in the ex-post harmonization efforts reviewed above provide every reason to believe that greater involvement of researchers in ex ante harmonization will lead to successful and useful harmonized data and SHARE is the best example yet of the value of researcher driven ex ante data collection efforts. Despite the shortcomings of past efforts, the potential harmonized data hold to help identify key relationships between policies, socio-economic factors, and health outcomes makes the money invested in them a worthwhile venture.

References

Börsch-Supan, Axel, Brugiavini, Agar, Jürges, Hendrik Mackenbach, Johan, Siegrist, Johannes, and Weber, Guglielmo. Health, Ageing and Retirement in Europe – First Results from the Survey of Health, Ageing and Retirement in Europe Eds., Mannheim: MEA, 2005.

Burkhauser, Richard V., 2001. What policymakers need to know about poverty dynamics. *Journal of Policy Analysis and Management*, 20 (4) (Fall): 757-759.

Burkhauser, Richard V., Butrica, Barbara A, Daly, Mary C. and Lillard, Dean R. 2001. The Cross-National Equivalent File: A product of cross-national research. In Irene Becker, Notburga Ott and Gabriele Rolf (eds.) *Soziale Sicherung in Einer Dynamischen Gesellschaft*. Frankfurt, Germany, Campus Verlagi. pp. 354-376.

Burkhauser, Richard V., Duncan, Greg J., Hauser, Richard and Berntsen, Roland. 1991 Wife or Frau, women do worse: A comparison of men and women in the United States and Germany following marital dissolution. *Demography*, 28(3) (August): 353-360.

Burkhauser, Richard V., Giles, Philip, Lillard, Dean R. and Schwarze, Johannes. Forthcoming. After death do us part: An analysis of the economic well being of widows in four countries. *Journal of Gerontology*.

Burkhauser, Richard V. and Smeeding, Timothy M. 2001. The role of micro-level panel data in policy research. *Schmollers Jahrbuch: Journal of Applied Social Science Studies, 121* (4): 469-500.

Deaton, Angus. 2003. Health, inequality, and economic development. *Journal of Economic Literature*. 41(1):

Deaton, Angus, and Paxson, Christina. 2001. Mortality, income, and income inequality over time in Britain and the United States.. National Bureau of Economic Research Working Paper 8534.

Duncan, Greg J., Gustafsson, Bjorn, Hauser, Richard, Schmauss, Gunther, Messinger, Hans, Muffels, Ruud, Nolan, Brian and Ray, Jean-Claude. 1993. Poverty Dynamics in Eight Countries. *Journal of Population Economics,* 6: 215-234.

Economic Outlook 2001. Organization for Economic Cooperation and Development, Paris, France.

Economic Outlook 2002. Organization for Economic Cooperation and Development, Paris, France.

Gottschalk, Peter, and Smeeding, Timothy M. 1997 Cross-national comparisons of earnings and income inequality. *Journal of Economic Literature*, 35(2): 633-687.

Gruber, Jonathan and Wise, David A. (eds.) 1999. *Social Security and Retirement around the World.* The University of Chicago Press for the NBER, Chicago.

Gruber, Jonathan and Wise, David A. (eds.) 2004. *Social Security Programs and Retirement Around the World: Micro-Simulation*. The University of Chicago Press for the NBER, Chicago.

Jenkins, Stephen, Schluter, Christian, and Wagner, Gert G. 2003. The dynamics of child poverty: Britain and Germany compared. *Journal of Comparative Family Studies* Summer, 34: 337-353.

Kenkel, D., Lillard, Dean. R., and Mathios, Alan. 2004. 'A Cross-Country Analysis of Tobacco Control Policies and Smoking Over the Life-course,' Journal d'Economie Medicale 22(3).

Lillard, Dean R. and Burkhauser, Richard V. 2005. Income inequality and health: A cross-country analysis. *Schmollers Jahrbuch: Journal of Applied Social Science Studies.*

Smeeding, Timothy M. 2004. Twenty years of research on income inequality, poverty, and redistribution in the developed world" *Socio-Economic Review*, 2: 149-163.

Wagstaff, Adam, and van Doorslaer, Eddy. 2000. Income inequality and health: What does the literature tell us? *Annual Review of Public Health*. 21: 543-567.

Appendix

This appendix describes the major micro-based data sets used in cross-national research. We also list the web addresses where researchers can find additional information, including information on how access each data set.

Luxembourg Income Study (LIS)

Web address: www.lisproject.org

Description: The Luxembourg Income Study (LIS) is a non-profit cooperative research project with a membership that includes 25 countries. The LIS project began in 1983. It is mainly funded by the national science and social science research foundations of its member countries. LIS now covers 29 countries with datasets that span up to three decades.

The LIS database is a collection of household income surveys. These surveys provide demographic, income and expenditure information on three different levels: household, person, and child. LIS's primary goal is to construct a harmonized database that is the best source for international comparative studies.

Type of underlying surveys: Cross-sectional.

Countries: Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Ireland, Israel, Italy, Luxembourg, Mexico, Netherlands, Norway, Poland, Romania, Russia, Slovak Republic, Slovenia, Spain, Sweden, Taiwan, Switzerland, United Kingdom, United States

Years represented in data: 1969-2000

Most cross-section years available for a single country: 9 years (Canada)

Cost: Use of data by electronic mail is free to all LIS Member Country researchers. All students may use the LIS data at no charge. However, researchers in non-member countries or non-subscribing international organizations will be assessed user fees. Use of the data is restricted to social science research purposes only. No private or commercial use is permitted.

Cross-National Equivalent File (CNEF)

Web address: www.human.cornell.edu/pam/gsoep/equivfil.cfm

Description: The Cross-National Equivalent File 1980-2002 contains equivalently defined variables for the Panel Study of Income Dynamics (PSID), the German Socio-Economic Panel (GSOEP), the British Household Panel Study (BHPS), and the Canadian Survey of Labour and Income Dynamics (SLID). The data are designed to allow cross-national researchers not experienced in panel data analysis to access a simplified version of these panels, while providing experienced panel data users with guidelines for formulating equivalent variables across

countries. Most importantly, the equivalent file provides a set of constructed variables (for example pre- and post-government income and United States and international household equivalence weights) that are not directly available on the original surveys. Since the Cross-National Equivalent File 1980-2002 can be merged with the original surveys, PSID-CNEF users can easily incorporate these constructed variables into current analyses.

Type of underlying surveys: All data are longitudinal. Canadian data are six-year rotating panels.

Countries: Canada (1992-2002), Germany (1984-2003), United Kingdom (1991-2002), United States (1980-1997, 1999, 2001, 2003),

Years represented in data: 1980-2003

Longest time span available for single country's data: 21 years (Germany and U.S.)

Cost: $125 for first time users, $30 for updates

Consortium of Household Panels for European Socio-economic Research (CHER)

Web address: www.ceps.lu/Cher/acceuil.cfm

Description: CHER contains comparable variables transformed according to a common plan and will be built by using standardized international classifications where available. Information in these files is available (a) for households and individuals on the micro level, (b) for single years and (c) as longitudinal information, all of them linked to macro and institutional data. The comparative database contains harmonized and consistent variables and identical data structures for each country included: 14 EU countries, Poland, Hungary, Canada and USA. The data are stored as system files for the statistical packages SPSS, SAS and Stata. They contain identical variable names, labels, values and data structures. Each country file is anonymized to be rated as a scientific use file. CHER data are available on a CD-ROM and are distributed to the scientific community, under rules for confidentiality and data protection.

Type of underlying surveys: Mostly panel (data from Sweden are cross-sectional).

Countries: Belgium (1992-1998), Germany (1990-2000), Hungary (1992-1997), Luxembourg (1995-2001), Poland (two non comparable panels 1994-1996 and 1997-2000), Switzerland (1999-2000), UK (1991-2001), USA (1990-1992).

Programs (but no data) are also available to convert data from the European Community Household Panel to be comparable to CHER data. The programs can be used to convert ECHP data for Denmark, France, Greece, Netherlands, Ireland, Italy, Portugal, and Spain (all for 1994-2001), Austria (1995-2000), and Finland (1996-2001). Cross-sectional data for Sweden are available for 1997-2000.

Years represented in data: 1990-2001

Longest time span available for single country's data: 11 years

Cost: CHER data subject to terms of individual country panels (some nominal fee may be assessed). Full panel of ECHP costs 8,000 Euros

European Community Household Panel (ECHP)

Web address: epunet.essex.ac.uk/echp.php (from the European Panel Users Network)

Description: The ECHP is a harmonized cross-national longitudinal survey focusing on household income and living conditions. It also includes items on health, education, housing, migration, demographics and employment characteristics. The ECHP is a valuable resource for researchers because it covers topics from multiple dimensions, data are comparable across countries, and a panel of (up to) eight years is available for each country.

Countries: Belgium, Denmark, Germany, Greece, Spain, France, Italy, Ireland, Luxembourg, The Netherlands, Portugal, the United-Kingdom (1994-2001). Austria (1995-2001), Finland (1996-2001). Sweden (1997-2001).

Type of underlying surveys: longitudinal except Sweden (which provides cross-sectional data derived from its National Survey on Living conditions).

Years represented in data: 1994-2001

Longest time span available for single country's data: 8 years

Cost: Full panel costs 8,000 Euros

European Union Survey of Living Conditions (EU-SILC)

Web address: www.cso.ie/eusilc/about_eusilc.htm

Description: The Survey on Income and Living Conditions (SILC) is an annual survey conducted by the central statistics offices of European Union countries. Its main purpose is to obtain information on the income and living conditions of different types of households. The survey also collects information on poverty and social exclusion. A representative random sample of households is surveyed.

This survey is being conducted throughout the European Union. The European Council and the Commission consider it as a major instrument in the fight against poverty and social exclusion. The European Union requires comparable and timely statistics to monitor this process.

Type of underlying surveys: Both longitudinal and cross-sectional but the eventual size of longitudinal samples is not yet clear.

Countries: Belgium, Czech Republic, Denmark, Germany, Estonia, Greece, Spain, France, Ireland, Italy, Cyprus, Latvia, Lithuania, Luxembourg, Hungary, Malta, Netherlands, Austria, Poland, Portugal, Slovakia, Slovenia, Finland, Sweden, United Kingdom, Iceland, Norway, Switzerland, Bulgaria, Croatia, Romania, Turkey.

Surveys were launched in 2003 in Belgium, Denmark, Greece, Ireland, Luxembourg, Austria, and Norway. They will be re-launched in 2004 in these countries, and in Spain, France, Italy, Portugal, Finland, Sweden, Iceland, Norway, Turkey, and Estonia. The SILC survey will be launched in 2005 in Germany, Netherlands, United Kingdom, Bulgaria, and Romania. Switzerland is scheduled to begin collecting data by 2007.

Data from Bulgaria, Croatia, Romania, and Turkey will not be fully comparable due to differences in the underlying data sources.

Longest time span available for single country's data: 2 years

Cost: Unknown

Survey of Health, Aging, and Retirement in Europe (SHARE)

Web address: www.share-project.org

Description: SHARE contains a set of 11 ex ante harmonized European country based data sets, each of which follows a cohort of older men and women. The questions and survey design are based on the U.S. Health and Retirement Study (HRS) and the English Longitudinal Study of Ageing (ELSA).Compared to single country data like the HRS and ELSA, SHARE allows researchers the advantage of capturing cross-national variation in the public policies, cultures and histories of a variety of European countries.

Data collected include health variables (e.g. self-reported health, physical functioning, cognitive functioning, health behaviour, use of health care facilities), psychological variables (e.g. psychological health, well-being, life satisfaction), economic variables (e.g. current work activity, job characteristics, opportunities to work past retirement age, sources and composition of current income, wealth and consumption, housing, education), and social support variables (e.g. assistance within families, transfers of income and assets, social networks, volunteer activities).

Type of underlying surveys: Planned to be longitudinal.

Countries: Austria, Belgium, Denmark, France, Germany, Greece, Italy, The Netherlands, Spain, Sweden, Switzerland. Further data are currently being collected in Israel.

Years represented in data: 2004 now available.

Longest time span available for single country's data: 1 year.

Table 1. The Distribution of Major Cross-National Data Harmonization Projects Across Policy Subject Areas

Cross-National Data	Income	Labor Outcomes	Health	Wealth
Cross Sectional	LIS EU-SILC?	LIS EU-SILC?	EU-SILC?	LIS?
Pre-Existing Panels	CNEF CHER	CNEF CHER	CNEF	
New Panels	ECHP EU-SILC? SHARE	ECHP EU-SILC? SHARE	ECHP SHARE	SHARE

Note: These are the major harmonized data projects. Some are reliable and some have the potential to be. A question mark indicates the potential for being a reliable harmonized data set.

CHER is Consortium of Household Panels for European Socio-Economic Research
CNEF is Cross-National Equivalent File
ECHP is European Community Household Panel
EU-SILC is Statistics on Income and Living Conditions
LIS is Luxembourg Income Study
SHARE is Survey of Health, Aging and Retirement in Europe